

Analysis for investment strategy in Athens Greece

Thomas Melistas

June 2019

Contents

1	Introduction/Business Problem	1
1.1	General introduction	1
1.2	Description of the goal	1
2	Data used and their processing	2
2.1	Data source	2
2.2	Processing the data	2
3	Methodology and exploratory data analysis	2
4	Results and Discussion	9
5	Conclusion	10

1 Introduction/Business Problem

1.1 General introduction

Greece is rapidly overcoming its financial crisis as financial and economic indices show. Tourism as well as trade equip its capital, Athens, with a large potential for growth and this potential has attracted many investors from all over the world. Investment companies are currently buying real estate and new shopping malls and business parks are on their plans.

1.2 Description of the goal

In this report we will use data in order to analyze which venues overcrowd the center of Athens and thus determine which investments will face high competition. Moreover, we provide certain dense areas with lack of specific venues. Of course, this will be valuable for investment agencies since an accurate report will provide information on which venues they should avoid spending money on. Our main assumption is that density of the same type of

business is unattractive for investing whereas overall density with lack of particular kinds of venues is considered a profitable way to fill the gap in the market.

2 Data used and their processing

2.1 Data source

For my report, data from the Foursquare API was used. The data contains the venues around the center of Athens together with among others their exact location, address and distance from the city center.

2.2 Processing the data

Since I was not interested in a particular venue, but rather in the category that each venue falls into, I obtained the category of each one. In the original data, there was information not relevant to our analysis such as the city, the country, the state and the cross street for each venue. The wise thing to do was to get rid of any information that would not be used in our analysis as this proves to be less confusing and requires less computational power to handle the data.

3 Methodology and exploratory data analysis

We will use various methods to analyze the data we obtained and get a better intuition regarding decisions stakeholders will make in the future. At first we visualize the data by creating a folium map and then a heatmap to better understand the density of venues. Following this, we divide our venues into categories (according to type and later according to distance from the center) and we find the ones overcrowding the city center. Those are the ones strategic investors should avoid investing money on due to high competition and lack of large potential share in their respective markets. Since overall density of an area is helpful for the growth of a business we explore which venues have little or no presence in the highly dense areas. Those are considered very profitable investments. We demonstrate risky investments by the use of a HISTOGRAM.

Moreover, various techniques of classification and prediction are useful to forecast whether certain types of investments are good investments or follow the investment pattern in Athens. We use KNN (K nearest Neighbors) and decision trees to facilitate our analysis and make accurate predictions.

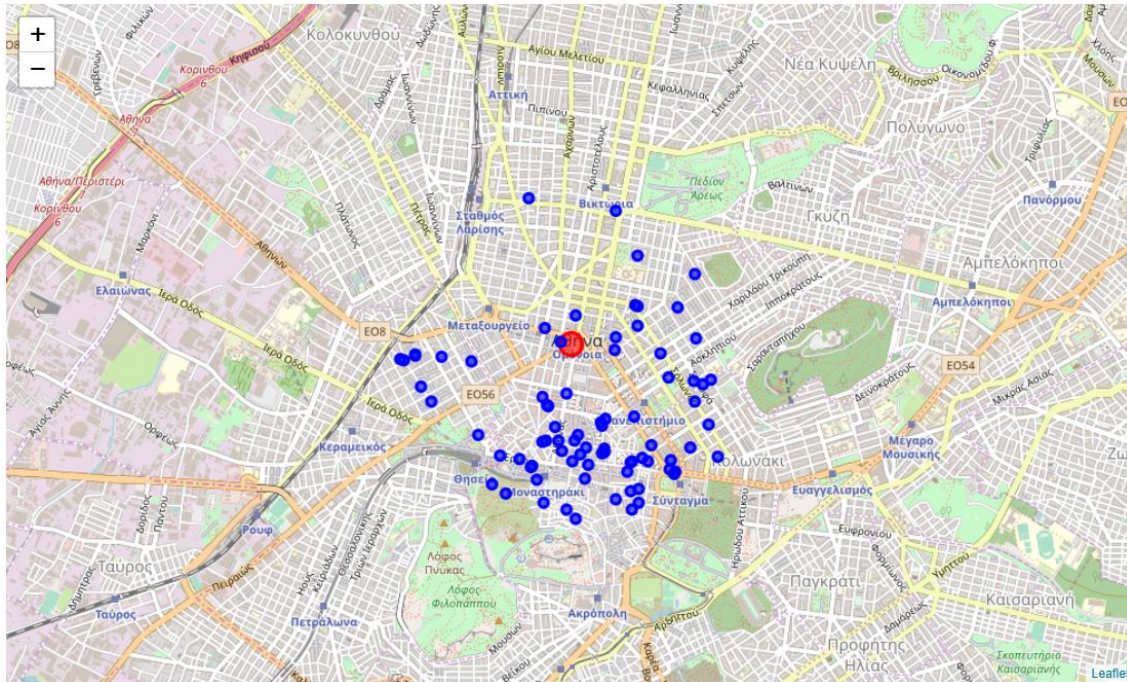
Let's have a first look to our data.

	name	categories	address	cc	city	country	crossStreet	distance	formattedAddress	labeledLatLngs	lat	lng	neighb
0	National Theatre of Greece (Εθνικό Θέατρο)	Theater	Αγίου Κωνσταντίνου 22-24	GR	Αθήνα	Ελλάδα	NaN	203	[Αγίου Κωνσταντίνου 22-24, 104 37 Αθήνα, Αττικ...	[[{"label": "display", "lat": 37.98474217257332...	37.984742	23.725541	
1	Στάνη Γαλακτοπωλείο	Dessert Shop	Μαρίκας Κοτοπούλη 10	GR	Αθήνα	Ελλάδα	Βερανζέρου	186	[Μαρίκας Κοτοπούλη 10 (Βερανζέρου), 104 32 Αθή...	[[{"label": "display", "lat": 37.98546282764882...	37.985463	23.727908	Metax
2	Loukoumades Ktistakis (Λουκουμάδες Χανίων Θ. Κ...	Donut Shop	Σωκράτους 59	GR	Αθήνα	Ελλάδα	NaN	69	[Σωκράτους 59, 104 31 Αθήνα, Αττική, Ελλάδα]	[[{"label": "display", "lat": 37.98393592722866...	37.983936	23.726768	
3	Της Θεάτρου Το Στέκι	Kafenio	7 Theatrou St	GR	Αθήνα	Ελλάδα	NaN	393	[7 Theatrou St, Αθήνα, Αττική, Ελλάδα]	[[{"label": "display", "lat": 37.98070281296481...	37.980703	23.725394	
4	Τράτα	Fish Taverna	Νικηταρά 9	GR	Αθήνα	Ελλάδα	Θεμιστοκλέους 8	301	[Νικηταρά 9 (Θεμιστοκλέους 8), 106 78 Αθήνα, Α...	[[{"label": "display", "lat": 37.98421877664627...	37.984219	23.730939	Athen
5	Taf Coffee	Coffee Shop	Εμμανουήλ Μπενάκη 7	GR	Αθήνα	Ελλάδα	NaN	293	[Εμμανουήλ Μπενάκη 7, 106 78 Αθήνα, Αττική, Ελ...	[[{"label": "display", "lat": 37.98345088308742...	37.983451	23.730853	
6	Καραμανλίδικα του Φάνη	Meze Restaurant	Σωκράτους 1	GR	Αθήνα	Ελλάδα	Ευριπίδου 52	434	[Σωκράτους 1 (Ευριπίδου 52), 105 54 Αθήνα, Αττ...	[[{"label": "display", "lat": 37.98016872811729...	37.980169	23.725757	

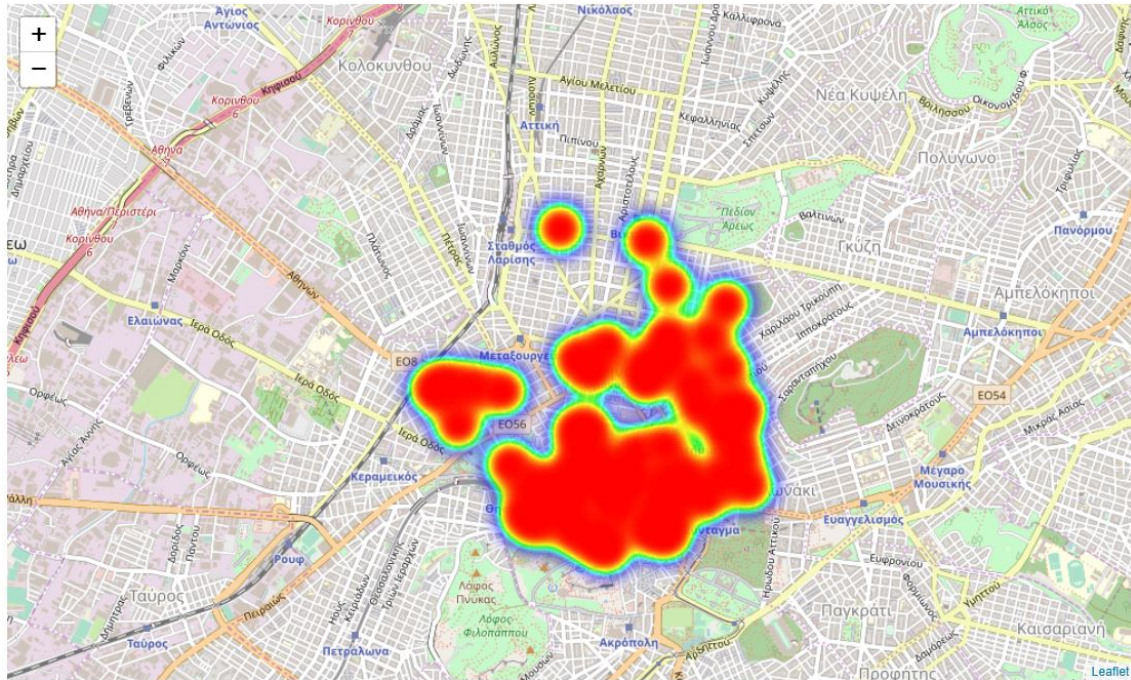
As we see, there is a lot of extra and unnecessary information so we clean our data. Our new dataframe looks like this.

	name	categories	address	distance	labeledLatLngs	lat	lng	neighborhood	postalCode	id
0	National Theatre of Greece (Εθνικό Θέατρο)	Theater	Αγίου Κωνσταντίνου 22-24	203	[{"label": "display", "lat": 37.98474217257332...	37.984742	23.725541	NaN	104 37	4e822369f5b91d2e46d16f27
1	Στάνη Γαλακτοπωλείο	Dessert Shop	Μαρίκας Κοτοπούλη 10	186	[{"label": "display", "lat": 37.9854632764882...	37.985463	23.727908	Metaxourgeio	104 32	4c038ca13f03b713b6fb5141
2	Loukoumades Ktistakis (Λουκουμάδες Χανίων Θ. Κ...	Donut Shop	Σωκράτους 59	69	[{"label": "display", "lat": 37.98393592722866...	37.983936	23.726768	NaN	104 31	4d4bdade1ae43704f77ced60
3	Της Θεάτρου Το Στίκι	Kafenio	7 Theatrou St	393	[{"label": "display", "lat": 37.98070281296481...	37.980703	23.725394	NaN	NaN	55156e65498eb8bdba7fd799
4	Τράτα	Fish Taverna	Νικηταρά 9	301	[{"label": "display", "lat": 37.98421877664627...	37.984219	23.730939	Omonia, Athens, Attica	106 78	4df4a6d218a88611c6c0c27e
5	Taf Coffee	Coffee Shop	Εμμανουήλ Μπενάκη 7	293	[{"label": "display", "lat": 37.98345088308742...	37.983451	23.730853	NaN	106 78	4b64103df964a520759d2ae3

It certainly looks cleaner and more concise. Now, let's have a spatial visualization of our data.



With the red dot, we can see the center of Athens, Greece. The blue dots represent the locations of the venues obtained by the Foursquare API. As we can see, there is high density south and southeast of the center. The better tool to make this more rigorous is a heat map based on our data.



It is clear that the north and northwest regions are less crowded. This is primarily due to the fact that the Acropolis is directly south of the red cloud, thus so far proximity to Acropolis seems to be the decisive factor for starting a business in the city center. As the cloud expands, it is natural to assume that it will expand in a concentric way around the city center depicted in red in the first map.

The first result is that investors should try to start filling this disk staying as close to the red cloud as possible. Less dense areas are the ones with lower competition thus easier to expand on. Whatever venue one will invest in there, it is likely to be successful. Furthermore, We should also provide an analysis on which venues one should focus on investing in the red cloud.

The following table provides all categories of venues in the are together with their number.

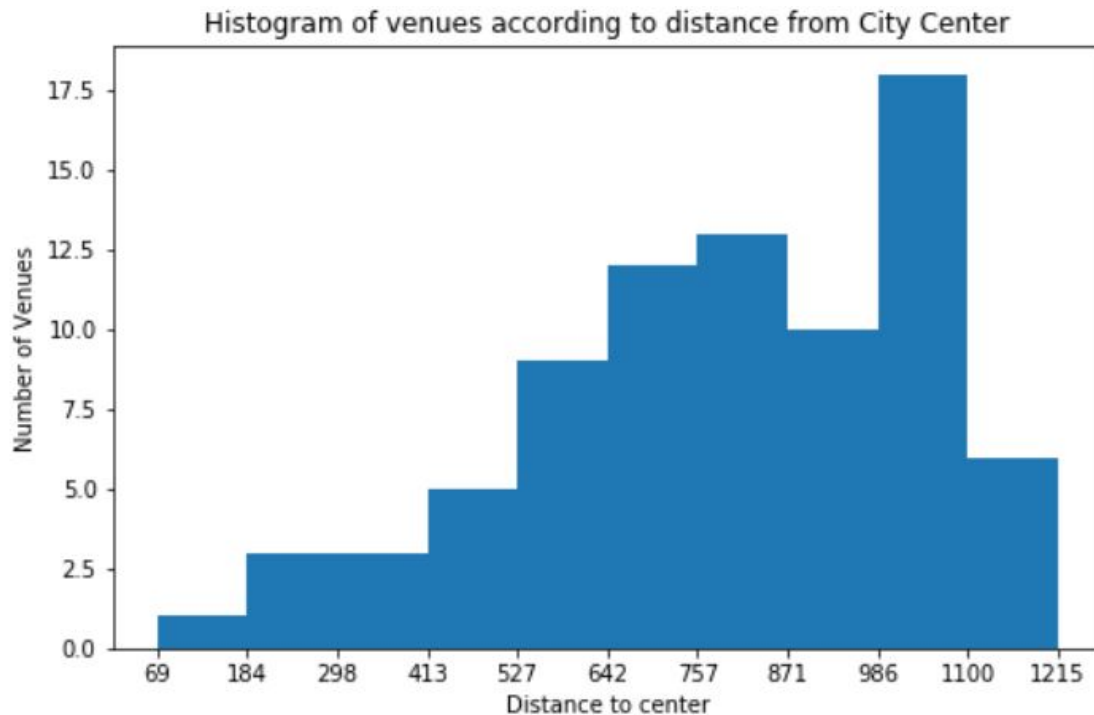
categories	Number
Art Gallery	1
Bar	7
Bistro	1
Bookstore	3
Boutique	1
Cafe	5
Chocolate Shop	1
Cocktail Bar	6
Coffee Shop	7
Dessert Shop	4
Donut Shop	1
Electronics Store	1
Falafel Restaurant	2
Fast Food Restaurant	1
Fish Taverna	1
Frozen Yogurt Shop	1
Gourmet Shop	2
Greek Restaurant	3
Historic Site	4
History Museum	1
Hotel	3
Ice Cream Shop	1
Indian Restaurant	1
Jewelry Store	1
Kafenio	1
Meze Restaurant	2
Middle Eastern Restaurant	1
Modern Greek Restaurant	1
Molecular Gastronomy Restaurant	1
Pizza Place	1
Plaza	1
Roof Deck	1
Souvlaki Shop	1
Spa	1
Sushi Restaurant	2
Theater	3
Whisky Bar	1
Wine Bar	4

Any category with less than 4 items is considered to have minor competition, so we will keep only the categories with at least 4 items. Those are demonstrated in the following dataframe.

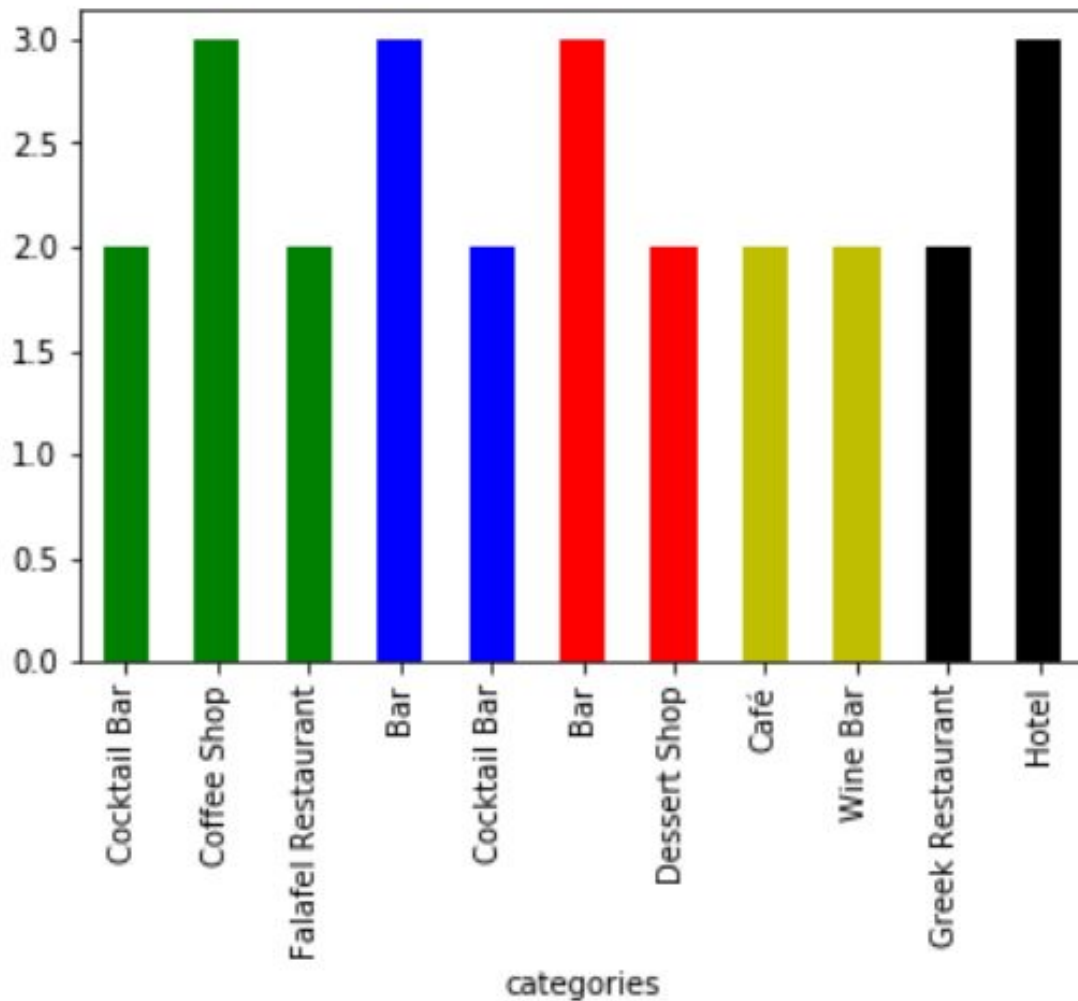
categories	Number
Bar	7
Café	5
Cocktail Bar	6
Coffee Shop	7
Dessert Shop	4
Wine Bar	4

The second result is that investors should avoid investing on venues of the categories in the above dataframe. If they have investment expertise in the above categories, we will try to find at which distance from the center those are located and find less crowded areas.

First of all, let's divide our venues according to their distance from the center. The following histogram shows their distribution.



It is evident that there is a gap in density between 871 and 985 meters from the city center, so this area is a good opportunity for investing in. The center is less crowded due to mainly operating costs for businesses, so the gap suggested above seems an optimal distance. More specifically, we are interested in exploring the location of the categories in the above dataframe. We divide them according to their distance from the center and we create a histogram encoding the ones in the same distance bin by the same color. Green is between 527 and 642 meters. Blue is between 642 and 757. Red is between 757 and 871. Yellow is between 871 and 986 and finally black is between 986 and 1100.



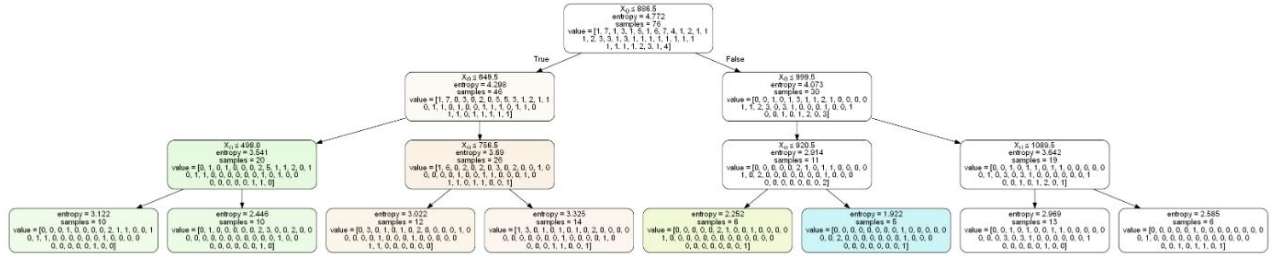
The third result is that any category not contained in the above colors of the histogram is permitted. For instance we see that there is no dessert shop and hotel in the green group so this might be a good idea for investment. (i.e if the stakeholders want to invest in a hotel, they'd better stay close to the city center)

Now is the time to perform our predictive analysis. We will use decision trees and the K nearest neighbors model (KNN). The input for the model will be the intentions of the investors with respect to the distance from the city center. If the investors intentions meet the overall pattern of the present venues then the investor should definitely avoid investing on his/her idea since competition will be high and the margin for profit low.

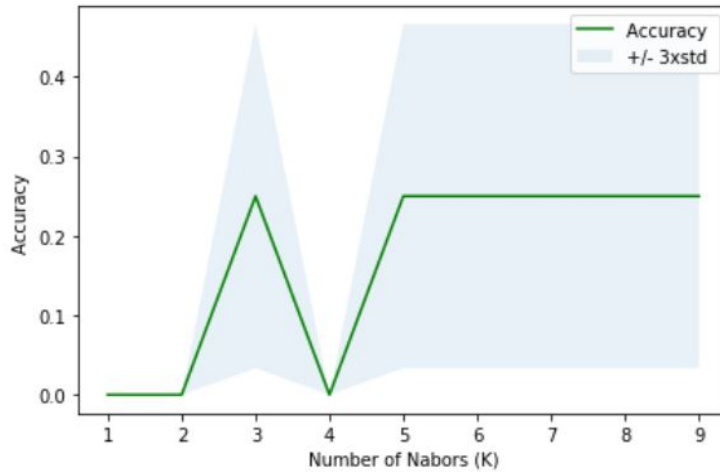
As we see since the data we have in our disposition is very limited, we cannot achieve high accuracy, thus our prediction models do not provide enough information on particular types of venues. A possible solution to this would be to collect more data about operating costs, profit, criminal records and build more advanced prediction models.

Let's start with the decision tree. In the following we can see that at each step the entropy decreases, but not by much, which causes our model to be with accuracy level 0.25.

This is pretty inaccurate and should not be used as a suggestive model.



Another alternative is KNN (K nearest neighbors). Of course, when dealing with KNN, we have to make sure that we use the optimal k. The following graphs shows the optimal ones.



As we can see $k = 3$ is the smallest one and achieves the highest accuracy level. This level is again unfortunately 0.25 which again indicates that the data we use are pretty restrictive and weak. Googlemaps API might have more and better information, yet it is not free and thus it is beyond the scope of this project.

Using SVM (support vector machine) is less enlightening since its f1 score and jaccard similarity score are close to 0, so we do not include this method in this report.

4 Results and Discussion

It is always very helpful to visualize our data and use all available methods in order to extract all of the information from our data set. The folium map gave us a general idea on how the venues are spread across the city center of Athens. We noticed that most venues are

located south and southeast. The heatmap confirmed this and added more intuition on the level of competition the future business is going to face. Sparsity is helpful among venues in the same category, yet it doesn't help much overall since the best idea is to invest in a dense area lacking the particular type of venue we are interested in building. For this reason, we divided our map into numerical categories based on the distance from the city center and we found the categories of venues with the most presence within each distance bin. This can be found in the second histogram above. Due to limitations coming from the data used, we were unable to extract any useful information when using predictive models such as KNN, SVM and Decision trees since the accuracy of our models was at most 0.25 thus this is the future direction of this project.

5 Conclusion

The scope of this project was to explore data for venues located in the center of Athens, Greece. By analyzing their density and location we were able to extract information about which investments will face high competition and thus will have high risk and low margin for profit and which are the ones that are considered good investments and highly profitable. Our report provides the answers to these questions.

The final decision on the optimal investment strategy will be made by investors based also on other characteristics of locations, taking into consideration additional factors like operating costs of each location (cost of rent or ownership), criminal rates, real estate availability, prices, social and economic dynamics of every neighborhood etc.

Future directions of this project is of course the effort to collect more data and build more accurate predictive classification models.